LAHIVE & COCKFIELD LLP

COUNSELLORS AT LAW
28 STATE STREET
BOSTON, MASSACHUSETTS 02109-1784
TELEPHONE (617) 227-7400
FAX (617) 742-4214
lc@lahive.com

Facsimile Cover Sheet

To: Examiner J. Seharaseyon
Company: U.S.P.T.O./Group 1647
Phone: 703-305-1112
Fax: 703-746-5177

From: Cynthia L. Kanik, Ph.D.
U.S.S.N. 09/554,933
Attorney Docket No: GIN-6713CPUS
Phone: (617) 227-7400
Fax: (617) 742-4214
Date: April 26, 2002
Pages (incl. cover): 64

Dear Examiner Seharaseyon,

Further to our previous conversation, I am faxing you a copy of the English translation of Priority Document No. 9-323129 filed on November 25, 1997. Please call me at (617) 227-7400 Ext. 251, if I can be of further assistance.

Sincerely,

Cynthia L. Kanik, Ph.D.

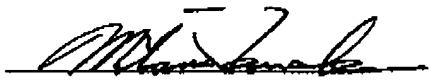
CERTIFICATE OF VERIFICATION

I, Mitsuo TANAKA

of c/o IMP Building 16th floor, 3-7, Shiromi 1-chome, Chuo-ku, Osaka 540-0001 Japan

state that the attached document is a true and complete translation to the best of my knowledge of Japanese Patent Application No. 9-323129 filed on November 25, 1997.

Dated this 12th day of April, 2002

Signature of translator: 

PATENT OFFICE

JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application:	November 25, 1997
Application Number:	Patent Application No. 323129/1997
Applicant:	SAGAMI CHEMICAL RESEARCH CENTER
	PROTEGENE INC.

Date:

Commissioner,

Patent Office

(Seal)

Document Name:	Application for Patent
Docket No.:	SO18128
Date of Application:	November 25, 1997
Addressee:	Commissioner, Patent Office
Title of the Invention:	Human Proteins Having Transmembrane Domains and DNAs Encoding These Proteins
Number of Claim(s):	6
Inventor:	
Address:	3-46-50, Wakamatsu, Sagamihara-shi, Kanagawa
Name:	Seishi KATO
Inventor:	
Address:	5-13-11, Takasago, Katsushika-ku, Tokyo
Name:	Tomoko YAMAGUCHI
Inventor:	
Address:	4-4-1, Nishi-Ohnuma, Sagamihara-shi, Kanagawa
Name:	Shingo SEKINE
Applicant:	
Representative Applicant:	
Identification No.:	000173762
Mail No.:	229
Address:	4-1, Nishi-Ohnuma 4-chome,

	Sagamihara-shi, Kanagawa	
Name:	SAGAMI CHEMICAL RESEARCH CENTER	
Representative:	Kiyoshi KONDO	
Telephone No.:	0427 (42) 4791	
Applicant:		
Identification No.:	596134998	
Mail No.:	153	
Address:	2-20-3, Naka-cho, Meguro-ku, Tokyo	
Name:	PROTEGENE INC.	
Representative:	Takeo TANAI	
Telephone No.:	03 (3792) 1019	
Payment of Fees:		
Payment Book No.:	011501	
Amount to be Paid:	¥21,000	
Attached Document:		
Item:	Specification	1 copy
Item:	Drawing	1 copy
Item:	Abstract	1 copy
Request for proof transmission:	Yes	

Document Name: Specification

Title of the Invention: Human Proteins Having Transmembrane Domains and DNAs Encoding These Proteins

Claim(s):

1. A protein comprising any one of the amino acid sequences represented by Sequence Nos. 1 to 3.

2. A DNA coding for the protein according to Claim 1.

3. A cDNA comprising any one of the base sequences represented by Sequence Nos. 4 to 6.

4. The cDNA according to Claim 3 consisting of any one of the base sequences represented by Sequence Nos. 7 to 9.

5. An expression vector capable of expressing the DNA according to any one of Claims 2 to 4 by in vitro translation or in eucaryotic cells.

6. A transformed eucaryotic cell capable of expressing the DNA according to any one of Claims 2 to 4 and of producing the protein according to Claim 1.

Detailed Explanation of the Invention:

[0001]

Art Field Related:

The present invention relates to human proteins having transmembrane domains, cDNAs coding for these proteins, and expression vectors of said cDNAs as well as

eucaryotic cells expressing said cDNAs. The proteins of the present invention can be employed as pharmaceuticals or as antigens for preparing antibodies against said proteins. The human cDNAs of the present invention can be utilized as probes for the gene diagnosis and gene sources for the gene therapy. Furthermore, the cDNAs can be utilized as gene sources for large-scale production of the proteins encoded by said cDNAs. Cells, wherein these membrane protein genes are introduced and membrane proteins are expressed in large amounts, can be utilized for detection of the corresponding ligands, screening of novel low-molecular pharmaceuticals, and so on.

[0002]

Prior Art:

Membrane proteins play important roles, as signal receptors, ion channels, transporters, etc. in the material transportation and the information transmission which are mediated by the cell membrane. Examples thereof include receptors for a variety of cytokines, ion channels for the sodium ion, the potassium ion, the chloride ion, etc., transporters for saccharides and amino acids, and so on, where the genes of many of them have been cloned already.

[0003]

It has been clarified that abnormalities of these membrane proteins are associated with a number of hitherto-

cryptogenic diseases. For instance, a gene of a membrane protein having twelve transmembrane domains was identified as the gene responsible for cystic fibrosis [Rommens, J. M. et al., Science 245: 1059-1065 (1989)]. In addition, it has been clarified that several membrane proteins act as receptors when a virus infects the cells. For instance, HIV-1 is revealed to infect into the cells through mediation of a membrane protein on the T-cell membrane, a CD-4 antigen, and a membrane protein having seven transmembrane domains, fusin [Feng, Y. et al., Science 272: 872-877 (1996)]. Therefore, discovery of a new membrane protein is anticipated to lead to elucidation of the causes of many diseases, so that isolation of a new gene coding for the membrane protein has been desired.

[0004]

Heretofore, owing to difficulty in the purification, many membrane proteins have been isolated by an approach from the gene side. A general method is the so-called expression cloning which comprises transfection of a cDNA library in eucaryotic cells to express cDNAs and then detection of the cells expressing the target membrane protein on the membrane by an immunological technique using an antibody or a physiological technique on the change in the membrane permeability. However, this method is applicable only to cloning of a gene of a membrane protein

with a known function.

[0005]

In general, membrane proteins possess hydrophobic transmembrane domains inside the proteins, wherein, after synthesis thereof in the ribosome, these domains remain in the phospholipid membrane to be trapped in the membrane. Accordingly, the evidence of the cDNA for encoding the membrane protein is provided by determination of the whole base sequence of a full-length cDNA followed by detection of highly hydrophobic transmembrane domains in the amino acid sequence of the protein encoded by said cDNA.

[0006]

Problems to be Solved by the Invention:

The object of the present invention is to provide novel human proteins having transmembrane domains, DNAs coding for said proteins, and expression vectors of said cDNAs as well as transformed eucaryotic cells that are capable of expressing said cDNAs.

[0007]

Means to Solve the Problems:

As the result of intensive studies, the present inventors have been successful in cloning of cDNAs coding for proteins having transmembrane domains from the human full-length cDNA bank, thereby completing the present invention. In other words, the present invention provides

human proteins having transmembrane domains, namely proteins containing any of the amino acid sequences represented by Sequence Nos. 1 to 3. Moreover, the present invention provides DNAs coding for the above-mentioned proteins, exemplified by cDNAs containing any of the base sequences represented by Sequence Nos. 4 to 9 as well as transformed eucaryotic cells that are capable of expressing said cDNAs.

[0008]

Mode for Carrying out the Invention

The proteins of the present invention can be obtained, for example, by a method for isolation from human organs, cell lines, etc., a method for preparation of peptides by the chemical synthesis based on the amino acid sequences of the present invention, or a method for production with the recombinant DNA technology using the DNAs coding for the transmembrane domains of the present invention, wherein the method for obtainment by the recombinant DNA technology is employed preferably. For instance, in vitro expression of the proteins can be achieved by preparation of an RNA by in vitro transcription from a vector having one of cDNAs of the present invention, followed by in vitro translation using this RNA as a template. Also, recombination of the translation region into a suitable expression vector by the method known in

the art leads to production of a large amount of the encoded protein by using prokaryotic cells such as *Escherichia coli*, *Bacillus subtilis*, etc., and eucaryotic cells such as yeasts, insect cells, mammalian cells, etc.

[0009]

In the case in which a protein of the present invention is produced by expression of one of the DNAs by in vitro translation, recombination of the translation region in said cDNA into a vector having an RNA polymerase promoter, followed by addition into an in vitro translation system such as a rabbit reticulocyte lysate, a wheat germ extract or the like, which contains an RNA polymerase corresponding to the promoter, allows in vitro production of the protein of the present invention. Examples of the RNA polymerase promoter include T7, T3, SP6, and so on. Vectors containing such an RNA polymerase promoter are exemplified by pKA1, pCDM8, pT3/T7 18, pT7/3 19, pBluescript II, and so on. Also, addition of the dog pancreas microsome etc. in the reaction system enables the membrane protein of the present invention to be expressed in a form integrated in the microsome membrane.

[0010]

In the case in which a protein of the present invention is produced by expression of a DNA in a microorganism such as *Escherichia coli* etc., a recombinant

expression vector bearing the translation region in the cDNA of the present invention is constructed in an expression vector having an origin, a promoter, a ribosome-binding site, a cDNA-cloning site, a terminator etc., which can be replicated in the microorganism, and, after transformation of the host cells with said expression vector, the thus-obtained transformant is incubated, whereby the protein encoded by said cDNA can be produced on a large scale in the microorganism. In this case, a protein fragment containing an optional region can be obtained by carrying out the expression with inserting an initiation codon and a termination codon in front of and behind an optional translation region. Alternatively, a fusion protein with another protein can be expressed. Only a protein portion encoded by said cDNA can be obtained by cleavage of said fusion protein with a suitable protease. Examples of the expression vector for *Escherichia coli* include the pUC system, pBluescript II, the pET expression system, the pGEX expression system, and so on.

[0011]

In the case in which one of the proteins of the present invention is produced by expression of a DNA in eucaryotic cells, the protein of the present invention can be produced as a membrane protein on the cell-membrane surface, when the translation region of said cDNA is

subjected to recombination to an expression vector for eucaryotic cells that has a promoter, a splicing region, a poly(A) insertion site, etc., followed by introduction into the eucaryotic cells. The expression vector is exemplified by pKA1, pCDM8, pSVK3, pMSG, pSVL, pBK-CMV, pBK-RSV, EBV vector, pRS, pYES2, and so on. Examples of eucaryotic cells to be used in general include mammalian culture cells such as simian kidney cells COS7, Chinese hamster ovary cells CHO, etc., budding yeasts, fission yeasts, silkworm cells, *Xenopus laevis* egg cells, and so on, but any eucaryotic cells may be used, provided that they are capable of expressing the present proteins on the membrane surface. The expression vector can be introduced in the eucaryotic cells by methods known in the art such as the electroporation method, the calcium phosphate method, the liposome method, the DEAE-dextran method, and so on.

[0012]

After one of the proteins of the present invention is expressed in prokaryotic cells or eucaryotic cells, the objective protein can be isolated from the culture and purified by a combination of separation procedures known in the art. Such examples include treatment with a denaturing agent such as urea or a surface-active agent, sonication, enzymatic digestion, salting-out or solvent precipitation, dialysis,

centrifugation, ultrafiltration, gel filtration, SDS-PAGE, isoelectric focusing, ion-exchange chromatography, hydrophobic chromatography, affinity chromatography, reverse phase chromatography, and so on.

[0013]

The proteins of the present invention include peptide fragments (5 amino acid residues or more) containing any partial amino acid sequence in the amino acid sequences represented by Sequence Nos. 1 to 3. These peptide fragments can be utilized as antigens for preparation of antibodies. Hereupon, among the proteins of the present invention, those having the signal sequence are secreted in the form of maturation proteins on the surface of the cells, after the signal sequences are removed. Therefore, these maturation proteins shall come within the scope of the protein of the present invention. The N-terminal amino acid sequences of the maturation proteins can be easily identified by using the method for the cleavage-site determination in a signal sequence [Japanese Patent Kokai Publication No. 1996-187100]. Furthermore, some membrane proteins undergo the processing on the cell surface to be converted to the secretory forms. Such proteins or peptides in the secretory forms shall come within the scope of the protein of the present invention. When sugar chain-binding sites are present in the amino

acid sequences, expression in appropriate eucaryotic cells affords proteins wherein sugar chains are added. Accordingly, such proteins or peptides wherein sugar chains are added shall come within the scope of the protein of the present invention.

[0014]

The DNAs of the present invention include all DNAs coding for the above-mentioned proteins. Said DNAs can be obtained by using a method by chemical synthesis, a method by cDNA cloning, and so on.

[0015]

The cDNAs of the present invention can be cloned, for example, from cDNA libraries of the human cell origin. These cDNA are synthesized by using as templates poly(A)$^+$ RNAs extracted from human cells. The human cells may be cells delivered from the human body, for example, by the operation or may be the culture cells. The cDNAs can be synthesized by using any method selected from the Okayama-Berg method [Okayama, H. and Berg, P., Mol. Cell. Biol. 2: 161-170 (1982)], the Gubler-Hoffman method [Gubler, U. and Hoffman, J. Gene 25: 263-269 (1983)], and so on, but it is preferred to use the capping method [Kato, S. et al., Gene 163: 193-196 (1995)], as exemplified in Examples, in order to obtain a full-length clone in an effective manner. In addition, commercially available, human cDNA libraries can

be utilized. Cloning of the cDNAs of the present invention from the cDNA libraries can be carried out by synthesis of an oligonucleotide on the basis of an optional portion in the cDNA base sequences of the present invention, followed by screening using this oligonucleotide as the probe according to the colony or plaque hybridization by a method known in the art. In addition, the cDNA fragments of the present invention can be prepared by synthesis of an oligonucleotide to be hybridized at both termini of the objective cDNA fragment, followed by the usage of this oligonucleotide as the primer for the RT-PCR method from an mRNA isolated from human cells.

[0016]

The cDNAs of the present invention are characterized by containing either of the base sequences represented by Sequence Nos. 4 to 6 or the base sequences represented by Sequence Nos. 7 to 9. Table 1 summarizes the clone number (HP number), the cells affording the cDNA clone, the total base number of the cDNA, and the number of the amino acid residues of the encoded protein, for each of the cDNAs.

[0017]

[Table 1]

Table 1

Sequence No.	HP No.	Cell	Number of bases	Number of amino acids
1, 4, 7	HP01207	Stomach Cancer	2938	269
2, 5, 8	HP01862	Stomach Cancer	2290	311
3, 6, 9	HP10493	PMA-U937	3705	383

[0018]

Hereupon, the same clones as the cDNAs of the present invention can be easily obtained by screening of the cDNA libraries constructed from the human cell lines and human tissues utilized in the present invention by the use of an oligonucleotide probe synthesized on the basis of the cDNA base sequence described in any of Sequence Nos. 4 to 9.

[0019]

In general, the polymorphism due to the individual difference is frequently observed in human genes. Accordingly, any cDNA that is subjected to insertion or deletion of one or plural nucleotides and/or substitution with other nucleotides in Sequence Nos. 4 to 9 shall come within the scope of the present invention.

[0020]

In a similar manner, any protein that is formed by these modifications comprising insertion or deletion of one or plural amino acids and/or substitution with other amino acids shall come within the scope of the present invention, as far as the protein possesses the activity of any protein having the amino acid sequences represented by Sequence Nos. 1 to 3.

[0021]

The cDNAs of the present invention include cDNA fragments (10 bp or more) containing any partial base sequence in the base sequences represented by Sequence Nos. 4 to 6 or in the base sequences represented by Sequences No. 7 to 9. Also, DNA fragments consisting of a sense chain and an anti-sense chain shall come within this scope. These DNA fragments can be utilized as the probes for the gene diagnosis.

[0022]

Examples

The present invention is embodied in more detail by the following examples, but this embodiment is not intended to restrict the present invention. The basic operations and the enzyme reactions with regard to the DNA recombination are carried out according to the literature ["Molecular Cloning. A Laboratory Manual" , Cold Spring Harbor Laboratory, 1989]. Unless otherwise stated,

restrictive enzymes and a variety of modification enzymes to be used were those available from TAKARA SHUZO. The manufacturer's instructions were used for the buffer compositions as well as for the reaction conditions, in each of the enzyme reactions. The cDNA synthesis was carried out according to the literature [Kato, S. et al., Gene 150: 243-250 (1994)].

[0023]

(1) Preparation of Poly(A)$^+$ RNA

The histiocyte lymphoma cell line U937 (ATCC CRL 1593) stimulated by phorbol ester and tissues of stomach cancer delivered by the operation were used for human cells to extract mRNAs. The cell line was incubated by a conventional procedure.

[0024]

After about 1 g of the human cells was homogenized in 20 ml of a 5.5 M guanidium thiocyanate solution, a total mRNA was prepared according to the literature [Okayama, H. et al., "Method in Enzymology", Vol. 164, Academic Press, 1987]. This was subjected to chromatography on oligo(dT)-cellulose column washed with a 20 mM Tris-hydrochloride buffer solution (pH 7.6), 0.5 M NaCl, and 1 mM EDTA to obtain a poly(A)$^+$ RNA according to the above-described literature.

[0025]

(2) Construction of cDNA Library

Ten micrograms of the above-mentioned poly(A)$^+$ RNA were dissolved in a 100 mM Tris-hydrochloride buffer solution (pH 8), one unit of an RNase-free, bacterial alkaline phosphatase was added, and the reaction was run at 37 ℃ for one hour. After the reaction solution was subjected to phenol extraction, followed by ethanol precipitation, the resulting pellet was dissolved in a solution containing 50 mM sodium acetate (pH 6), 1 mM EDTA, 0.1% 2-mercaptoethanol, and 0.01% Triton X-100. Thereto was added one unit of a tobacco-origin acid pyrophosphatase (Epicentre Technologies) and a total 100 µl volume of the resulting mixture was reacted at 37℃ for one hour. After the reaction solution was subjected to phenol extraction, followed by ethanol precipitation, the resulting pellet was dissolved in water to obtain a solution of a decapped poly(A)$^+$ RNA.

[0026]

The decapped poly(A)$^+$ RNA and 3 nmol of a chimeric DNA-RNA oligonucleotide (5'-dG-dG-dG-dG-dA-dA-dT-dT-dC-dG-dA-G-G-A-3') were dissolved in an aqueous solution containing 50 mM Tris-hydrochloride buffer solution (pH 7.5), 0.5 mM ATP, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, and 25% polyethylene glycol, whereto was added 50 units of T4 RNA ligase and a total 30 µl volume of the resulting

mixture was reacted at 20 ℃ for 12 hours. After the reaction solution was subjected to phenol extraction, followed by ethanol precipitation, the resulting pellet was dissolved in water to obtain a chimeric-oligo-capped poly(A)$^+$ RNA.

[0027]

After digestion of vector pKA1 (Japanese Patent Kokai Publication No. 1992-117292) developed by the present inventors with KpnI, about 60 dT tails were added using a terminal transferase. A vector primer to be used below was prepared by digestion of this product with EcoRV to remove a dT tail at one side.

[0028]

After 6 μg of the previously-prepared chimeric-oligo-capped poly(A)$^+$ RNA was annealed with 1.2 μg of the vector primer, the resulting product was dissolved in a solution containing 50 mM Tris-hydrochloride buffer solution (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, and 1.25 mM dNTP (dATP + dCTP + dGTP + dTTP), 200 units of a reverse transcriptase (GIBCO-BRL) were added, and the reaction in a total 20 μl volume was run at 42℃ for one hour. After the reaction solution was subjected to phenol extraction, followed by ethanol precipitation, the resulting pellet was dissolved in a solution containing 50 mM Tris-hydrochloride buffer

solution (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol. Thereto were added 100 units of EcoRI and a total 20 µl volume of the resulting mixture was reacted at 37 ℃ for one hour. After the reaction solution was subjected to phenol extraction, followed by ethanol precipitation, the resulting pellet was dissolved in a solution containing 20 mM Tris-hydrochloride buffer solution (pH 7.5), 100 mM KCl, 4 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, and 50 µg/ml of the bovine serum albumin. Thereto were added 60 units of an *Escherichia coli* DNA ligase and the resulting mixture was reacted at 16℃ for 16 hours. To the reaction solution were added 2 µl of 2 mM dNTP, 4 units of *Escherichia coli* DNA polymerase I, and 0.1 unit of *Escherichia coli* RNase H and the resulting mixture was reacted at 12℃ for one hour and then at 22℃ for one hour.

[0029]

Next, the cDNA-synthesis reaction solution was used for transformation of *Escherichia coli* DH12S (GIBCO-BRL). The transformation was carried out by the electroporation method. A portion of the transformant was spread on the 2xYT agar culture medium containing 100 µg/ml ampicillin and the medium was incubated at 37℃ overnight. A colony formed on the agar medium was picked up at random and inoculated on 2 ml of the 2xYT culture medium containing 100 µg/ml ampicillin. After incubation at 37℃

overnight, the culture mixture was centrifuged to separate the cells, from which a plasmid DNA was prepared by the alkaline lysis method. The plasmid DNA was subjected to double digestion with EcoRI and NotI, followed by 0.8% agarose gel electrophoresis, to determine the size of the cDNA insert. Furthermore, using the thus-obtained plasmid as a template, the sequence reaction was carried out by using an M13 universal primer labeled with a fluorescent dye and a Taq polymerase (a kit of Applied Biosystems) and then the product was examined with a fluorescent DNA sequencer (Applied Biosystems) to determine an about 400-bp base sequence at the 5'-terminus of the cDNA. The sequence data were filed as the homo/protein cDNA bank database.

[0030]

(3) Selection of cDNAs Encoding Proteins Having Transmembrane Domains

A base sequence registered in the homo/protein cDNA bank was converted to three frames of amino acid sequences and the presence or absence of an open reading frame (ORF) beginning from the initiation codon was examined. Then, the selection was made for the presence of a signal sequence that is characteristic to a secretory protein at the N-terminus of the portion encoded by the ORF. These clones were sequenced from the both 5' and 3' directions by the use of the deletion method using

exonuclease III to determine the whole base sequence. The hydrophobicity/hydrophilicity profiles were obtained for proteins encoded by the ORF by the Kyte-Doolittle method [Kyte, J. & Doolittle, R. F., J. Mol. Biol. 157: 105-132 (1982)] to examine the presence or absence of a hydrophobic region. In the case in which there is a hydrophobic region of a putative transmembrane domain in the amino acid sequence of an encoded protein, this protein was judged as a membrane protein.

[0031]

(4) Functional Verification of Secretory Signal Sequence or Transmembrane Domains

It was verified by the method described in the literature [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)] that the N-terminal hydrophobic region in the secretory protein clone candidate obtained in the above-mentioned steps functions as a secretory signal sequence. First, the plasmid containing the target cDNA was cleaved at an appropriate restriction enzyme site existing at the downstream of the portion expected for encoding the secretory signal sequence. In the case in which this restriction site was a protruding terminus, the site was blunt-ended by the Klenow treatment or treatment with the mung-bean nuclease. Digestion with HindIII was further carried out and a DNA fragment containing the SV40 promoter

and a cDNA encoding the secretory signal sequence at the downstream of the promoter was separated by agarose gel electrophoresis. The resulting fragment was inserted between HindIII in pSSD3 (DDBJ/EMBL/GenBank Registration No. AB007632) and a restriction enzyme site selected so as to match with the urokinase-coding frame, thereby constructing a vector expressing a fusion protein of the secretory signal sequence of the target cDNA and the urokinase protease domain.

[0032]

After *Escherichia coli* (host: JM109) bearing the fusion-protein expression vector was incubated at 37℃ for 2 hours in 2 ml of the 2xYT culture medium containing 100 μg/ml of ampicillin, the helper phage M13KO7 (50 μl) was added and the incubation was continued at 37℃ overnight. A supernatant separated by centrifugation underwent precipitation with polyethylene glycol to obtain single-stranded phage particles. These particles were suspended in 100 μl of 1 mM Tris-0.1 mM EDTA, pH 8 (TE). Also, there were used as controls suspensions of single-stranded phage particles prepared in the same manner from pSSD3 and from the vector pKA1-UPA containing a full-length cDNA of urokinase [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)].

[0033]

The culture cells originating from the simian kidney, COS7, were incubated at 37℃ in the presence of 5% CO_2 in the Dulbecco's modified Eagle's culture medium (DMEM) containing 10% fetal calf serum. Into a 6-well plate (Nunc Inc., 3 cm in the well diameter) were inoculated 1×10^5 COS7 cells and incubation was carried out at 37℃ for 22 hours in the presence of 5% CO_2. After the culture medium was removed, the cell surface was washed with a phosphate buffer solution and then washed again with DMEM containing 50 mM Tris-hydrochloric acid (pH 7.5) (TDMEM). To the resulting cells was added a suspension of 1 μl of the single-stranded phage suspension, 0.6 ml of the DMEM culture medium, and 3 μl of TRANSFECTAM™ (IBF Inc.) and the resulting mixture was incubated at 37℃ for 3 hours in the presence of 5% CO_2. After the sample solution was removed, the cell surface was washed with TDMEM, 2 ml per well of DMEM containing 10% fetal calf serum was added, and the incubation was carried out at 37℃ for 2 days in the presence of 5% CO_2.

[0034]

To 10 ml of a 50 mM phosphate buffer solution (pH 7.4) containing 2% bovine fibrinogen (Miles Inc.), 0.5% agarose, and 1 mM calcium chloride were added 10 units of human thrombin (Mochida Pharmaceutical Co., Ltd.) and the resulting mixture was solidified in a plate of 9 cm in

diameter to prepare a fibrin plate. Ten microliters of the culture supernatant of the tansfected COS7 cells were spotted on the fibrin plate, which was incubated at 37℃ for 15 hours. In the case in which a clear circle appears on the fibrin plate, it is judged that the cDNA fragment codes for the amino acid sequence functioning as a secretory signal sequence. On the other hand, in case in which a clear circle is not formed, the cells were washed well, then the fibrin sheet was placed on the cells, and incubation was carried out at 37℃ for 15 hours. In case in which a clear portion is formed on the fibrin sheet, it indicates that the urokinase activity was expressed on the cell surface. In other words, the cDNA fragment is judged to code for the transmembrane domains.

[0035]

(5) Protein Synthesis by In Vitro Translation

The plasmid vector bearing the cDNA of the present invention was used for in vitro transcription/translation with a T_NT rabbit reticulocyte lysate kit (Promega). In this case, [^{35}S]methionine was added to label the expression product with a radioisotope. Each of the reactions was carried out according to the protocols attached to the kit. Two micrograms of the plasmid was reacted at 30℃ for 90 minutes in a total 25 μl volume of the reaction solution containing 12.5 μl of T_NT

rabbit reticulocyte lysate, 0.5 μl of a buffer solution (attached to kit), 2 μl of an amino acid mixture (methionine-free), 2 μl of [^{35}S]methionine (Amersham) (0.37 MBq/μl), 0.5 μl of T7 RNA polymerase, and 20 U of RNasin. To 3 μl of the resulting reaction solution was added 2 μl of the SDS sampling buffer (125 mM Tris-hydrochloric acid buffer, pH 6.8, 120 mM 2-mercaptoethanol, 2% SDS solution, 0.025% bromophenol blue, and 20% glycerol) and the resulting mixture was heated at 95℃ for 3 minutes and then subjected to SDS-polyacrylamide gel electrophoresis. The molecular weight of the translation product was determined by carrying out the autoradiograph.

[0036]

(6) Expression by COS7

Escherichia coli bearing the expression vector of the protein of the present invention was infected with helper phage M13KO7 and single-stranded phage particles were obtained by the above-mentioned procedure. The thus-obtained phage was used for introducing each expression vector in the culture cells originating from the simian kidney, COS7 by the above-mentioned procedure. After incubation at 37℃ for 2 days in the presence of 5% CO_2, the incubation was continued for one hour in the culture medium containing [^{35}S]cysteine or [^{35}S]methionine. Collection and lysis of the cells, followed by subjecting

to SDS-PAGE, allowed to observe the presence of a band corresponding to the expression product of each protein, which did not exist in the COS7 cells.

[0037]

(7) Clone Examples

<HP01207> (Sequence Nos. 1, 4, and 7)

Determination of the whole base sequence of the cDNA insert of clone HP01207 obtained from cDNA libraries of human stomach cancer revealed the structure consisting of a 100-bp 5'-nontranslation region, an 810-bp ORF, and a 2028-bp 3'-nontranslation region. The ORF codes for a protein consisting of 269 amino acid residues and there existed seven putative transmembrane domains. Figure 1 depicts the hydrophobicity/hydrophilicity profile, obtained by the Kyte-Doolittle method, of the present protein. In vitro translation resulted in formation of a smear translation product of a high molecular weight.

[0038]

The search of the protein data base by using the amino acid sequence of the present protein revealed that the protein was analogous to the mouse Surf-4 protein (PIR Accession No. A34727). Table 2 shows the comparison of the amino acid sequence between the human protein of the present invention (HP) and the mouse Surf-4 protein (MM). Therein, the marks of * and . represent an amino acid

residue identical with the protein of the present invention and an amino acid residue analogous to the protein of the present invention, respectively. The both proteins possessed a homology of 99.3% in the entire region.

[0039]

[Table 2]

Table 2

```
HS MGQNDLMGTAEDFADQFLRVTKQYLPHVARLCLISTFLEDGIRMWFQWSEQRDYIDTTWN
   ************************************************************.
MM MGQNDLMGTAEDFADQFLRVTKQYLPHVARLCLISTFLEDGIRMWFQWSEQRDYIDTTWS
HS CGYLLASSFVFLNLLGQLTGCVLVLSRNFVQYACFGLFGIIALQTIAYSILWDLKFLMRN
   ************************************************************
MM CGYLLASSFVFLNLLGQLTGCVLVLSRNFVQYACFGLFGIIALQTIAYSILWDLKFLMRN
HS LALGGGLLLLLAESRSEGKSMFAGVPTMRESSPKQYMQLGGRVLLVLMFMTLLHFDASFF
   ************************************************************
MM LALGGGLLLLLAESRSEGKSMFAGVPTMRESSPKQYMQLGGRVLLVLMFMTLLHFDASFF
HS SIVQNIVGTALMILVAIGFKTKLAALTLVVWLFAINVYFNAFWTIPVYKPMHDFLKYDFF
   **.*********************************************************
MM SIIQNIVGTALMILVAIGFKTKLAALTLVVWLFAINVYFNAFWTIPVYKPMHDFLKYDFF
HS QTMSVIGGLLLVVALGPGGVSMDEKKKEW
   ****************************
MM QTMSVIGGLLLVVALGPGGVSMDEKKKEW
```

[0040]

Furthermore, the search of the GenBank using the base sequences of the present cDNA has revealed the registration of a base sequence that exhibited an analogy of 98.6% with a 762-bp part from position 122 up to

position 883 (GenBank Accession No. Y14820), which codes for the fragment of the present protein.

[0041]

The mouse Surf-4 protein is one of proteins which are encoded in the mouse surfeit locus and has been considered to a housekeeping protein that is essential to the survival of cells [Huxley, C. et al., Mol. Cell. Biol. 10: 605-614 (1990)].

[0042]

<HP01862> (Sequence Nos. 2, 5 and 8)

Determination of the whole base sequence of the cDNA insert of clone HP01862 obtained from cDNA libraries of human stomach cancer revealed the structure consisting of an 80-bp 5'-nontranslation region, a 936-bp ORF, and a 1274-bp 3'-nontranslation region. The ORF codes for a protein consisting of 311 amino acid residues and there existed seven transmembrane domains. Figure 2 depicts the hydrophobicity/hydrophilicity profile, obtained by the Kyte-Doolittle method, of the present protein. In vitro translation resulted in formation of a smear translation product of a high molecular weight.

[0043]

The search of the protein data base using the amino acid sequence of the present protein has revealed the presence of sequences that were analogous to the rat NMDA

receptor glutamate-binding subunit (GenBank Accession No. S19586). Table 3 shows the comparison of the amino acid sequence between the human protein of the present invention (HP) and the rat NMDA receptor glutamate-binding subunit (RN). Therein, the marks of -, *, and . represent a gap, an amino acid residue identical with the protein of the present invention, and an amino acid residue analogous to the protein of the present invention, respectively. The both proteins possessed a homology of 41.0%.

[0044]

[Table 3]

Table 3

```
HS                                                    MSNPSAPPPYEDRNP

RN MKRVSWSLGTAILPQTLAILWGHKPLCLPMFSLPTLGPHTHRPLSSPLPMVNQGIPMVPV
HS LYPGPLPPGGYGQPSVLPGGYPAYPGYPQPGYGHPAGYPQPMPPTHPMPMNYGPGHGYDG
          **  .  .  .  . .* **. * .*.   **.*.  ...* .* *    .  **   **.
RN PITRWLPLKDLLKEATHQGHYPQSP-FPPNPYGQPPPFQDPGSPQHGNYQEEGPPSYYDN
HS EERAVSDSFGPGEWDDRKVRHTFIRKVYSIISVQLLITVAIIAIFTFVEPVSAFVRRNVA
   ..       * .  .* *...*..*****.  ....** .*....******.  *..*** **
RN QD------FPSVNW-DKSIRQAFIRKVFLVLTLQLSVTLSTVAIFTFVGEVKGFVRANVW
HS VYYVSYAVFVVTYLILACCQGPRRRFPWNIILLTLFTFAMGFMTGTISSMYQTKAVIIAM
   .******.* ..  ..*.** .  *.. ***..  *...*.....*.* *.*.*.*.***.*.
RN TYYVSYAIFFISLIVLSCCGDFRKKHPWNLVALSILTISLSYMVGMIASFYNTEAVIMAV
HS IITAVVSISVTIFCFQTKVDFTSCTGLFCVLGIVLLVTGIVTSIVLYFQYVYWLHMLYAA
    **..*...*.**..**.  ***** *.  *  .**..  .*.     ....     *...**.
RN GITTAVCFTVVIFSMQTRYDFTSCMGVLLVSVVVLFIFAIL---CIFIRNRI-LEIVYAS
HS LGAICFTLFLAYDTQLVLGNRKHTISPEDYITGALQIYTDIIYIFTFVLQLMGDRN
```

. ** *** *.***.. ..***.*. .**..***** ** ..* ..*
RN LGALLFTCFLAVDTQLLLGNKQLSLSPEEYVFAALNLYTDIINIFLYILTIIGRSQGIGQ

[0045]

Furthermore, the search of the GenBank using the base sequences of the present cDNA has revealed the presence of sequences that possessed a homology of 90% or more (for example, Accession No. H06014) in EST, but any of the sequences was shorter than the present cDNAs and was not found to contain the initiation codon.

[0046]

The rat NMDA receptor glutamate-binding subunit is one of subunits of an NMDA receptor complex which exist specifically in the brain [Kumar, K. N. et al., Nature 354: 70-73 (1991)]. The protein of the present invention has seven transmembrane domains characteristic to channels and transporters and thereby is considered to play a role as a channel and a transporter.

[0047]

<HP10493> (Sequence Nos. 3, 6 and 9)

Determination of the whole base sequence of the cDNA insert of clone HP10493 obtained from cDNA libraries of the human lymphoma U937 revealed the structure consisting of a 123-bp 5'-nontranslation region, a 1152-bp

ORF, and a 2430-bp 3'-nontranslation region. The ORF codes for a protein consisting of 383 amino acid residues and there existed one transmembrane domain at the N-terminus. Figure 3 depicts the hydrophobicity/hydrophilicity profile, obtained by the Kyte-Doolittle method, of the present protein. Introduction of an expression vector, wherein the HindIII-AccI fragment containing a cDNA portion coding for the N-terminal 44 amino acid residues of the present protein was inserted into the HindIII-PmaCI site of pSSD3, into the COS7 cells revealed the urokinase activity on the cell surface to indicate that the present protein is the type-II membrane protein. In vitro translation resulted in formation of a translation product of 43 kDa that was almost consistent with the molecular weight of 43,001 predicted from the ORF.

[0048]

The search of the protein data base using the amino acid sequence of the present protein has not revealed the presence of any known protein having an analogy. The search of the motif sequences has revealed a high probability that histidine at position 175 is an active site of the trypsin-type serine protease. Accordingly, the present protein is likely to be a membrane-type protease. Also, the search of the GenBank using the base sequences of the present cDNA has revealed the presence of sequences

that possessed a homology of 90% or more (for example, Accession No. R81003) in EST, but many sequences were not distinct and the same ORF as that in the present cDNA was not found.

[0049]

Effects of the Invention:

The present invention provides human proteins having transmembrane domains, cDNAs coding for these proteins, and expression vectors of said cDNAs as well as eucaryotic cells expressing said cDNAs. All of the proteins of the present invention exist in the cell membrane, so that they are considered to be proteins controlling the proliferation and the differentiation of the cells. Accordingly, the proteins of the present invention can be employed as pharmaceuticals such as carcinostatic agents relating to the control of the proliferation and the differentiation of the cells or as antigens for preparing antibodies against said proteins. The cDNAs of the present invention can be utilized as probes for the gene diagnosis and gene sources for the gene therapy. Furthermore, the cDNAs can be utilized for large-scale expression of said proteins. Cells, wherein these membrane protein genes are introduced and membrane proteins are expressed in large amounts, can be utilized for detection of the corresponding ligands, screening of novel low-molecular pharmaceuticals,

and so on.

[0050]

Sequence Listing:

SEQ ID NO: 1

LENGTH: 269

TYPE: Amino acid

TOPOLOGY: Linear

MOLECULE TYPE: Protein

HYPOTHETICAL: No

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Stomach cancer

CLONE: HP01207

SEQUENCE DESCRIPTION:

Met Gly Gln Asn Asp Leu Met Gly Thr Ala Glu Asp Phe Ala Asp Gln
1 5 10 15

Phe Leu Arg Val Thr Lys Gln Tyr Leu Pro His Val Ala Arg Leu Cys
20 25 30

Leu Ile Ser Thr Phe Leu Glu Asp Gly Ile Arg Met Trp Phe Gln Trp
35 40 45

Ser Glu Gln Arg Asp Tyr Ile Asp Thr Thr Trp Asn Cys Gly Tyr Leu
50 55 60

Leu Ala Ser Ser Phe Val Phe Leu Asn Leu Leu Gly Gln Leu Thr Gly

65 70 75 80

Cys Val Leu Val Leu Ser Arg Asn Phe Val Gln Tyr Ala Cys Phe Gly

85 90 95

Leu Phe Gly Ile Ile Ala Leu Gln Thr Ile Ala Tyr Ser Ile Leu Trp

100 105 110

Asp Leu Lys Phe Leu Met Arg Asn Leu Ala Leu Gly Gly Gly Leu Leu

115 120 125

Leu Leu Leu Ala Glu Ser Arg Ser Glu Gly Lys Ser Met Phe Ala Gly

130 135 140

Val Pro Thr Met Arg Glu Ser Ser Pro Lys Gln Tyr Met Gln Leu Gly

145 150 155 160

Gly Arg Val Leu Leu Val Leu Met Phe Met Thr Leu Leu His Phe Asp

165 170 175

Ala Ser Phe Phe Ser Ile Val Gln Asn Ile Val Gly Thr Ala Leu Met

180 185 190

Ile Leu Val Ala Ile Gly Phe Lys Thr Lys Leu Ala Ala Leu Thr Leu

195 200 205

Val Val Trp Leu Phe Ala Ile Asn Val Tyr Phe Asn Ala Phe Trp Thr

210 215 220

Ile Pro Val Tyr Lys Pro Met His Asp Phe Leu Lys Tyr Asp Phe Phe

225 230 235 240

Gln Thr Met Ser Val Ile Gly Gly Leu Leu Leu Val Val Ala Leu Gly

245 250 255

Pro Gly Gly Val Ser Met Asp Glu Lys Lys Lys Glu Trp

260 265

[0051]

SEQ ID NO: 2

LENGTH: 311

TYPE: Amino acid

TOPOLOGY: Linear

MOLECULE TYPE: Protein

HYPOTHETICAL: No

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Stomach cancer

CLONE: HP01862

SEQUENCE DESCRIPTION:

Met Ser Asn Pro Ser Ala Pro Pro Pro Tyr Glu Asp Arg Asn Pro Leu
1 5 10 15

Tyr Pro Gly Pro Leu Pro Pro Gly Gly Tyr Gly Gln Pro Ser Val Leu
20 25 30

Pro Gly Gly Tyr Pro Ala Tyr Pro Gly Tyr Pro Gln Pro Gly Tyr Gly
35 40 45

His Pro Ala Gly Tyr Pro Gln Pro Met Pro Pro Thr His Pro Met Pro
50 55 60

Met Asn Tyr Gly Pro Gly His Gly Tyr Asp Gly Glu Glu Arg Ala Val
65 70 75 80

Ser Asp Ser Phe Gly Pro Gly Glu Trp Asp Asp Arg Lys Val Arg His

85 90 95

Thr Phe Ile Arg Lys Val Tyr Ser Ile Ile Ser Val Gln Leu Leu Ile

100 105 110

Thr Val Ala Ile Ile Ala Ile Phe Thr Phe Val Glu Pro Val Ser Ala

115 120 125

Phe Val Arg Arg Asn Val Ala Val Tyr Tyr Val Ser Tyr Ala Val Phe

130 135 140

Val Val Thr Tyr Leu Ile Leu Ala Cys Cys Gln Gly Pro Arg Arg Arg

145 150 155 160

Phe Pro Trp Asn Ile Ile Leu Leu Thr Leu Phe Thr Phe Ala Met Gly

165 170 175

Phe Met Thr Gly Thr Ile Ser Ser Met Tyr Gln Thr Lys Ala Val Ile

180 185 190

Ile Ala Met Ile Ile Thr Ala Val Val Ser Ile Ser Val Thr Ile Phe

195 200 205

Cys Phe Gln Thr Lys Val Asp Phe Thr Ser Cys Thr Gly Leu Phe Cys

210 215 220

Val Leu Gly Ile Val Leu Leu Val Thr Gly Ile Val Thr Ser Ile Val

225 230 235 240

Leu Tyr Phe Gln Tyr Val Tyr Trp Leu His Met Leu Tyr Ala Ala Leu

245 250 255

Gly Ala Ile Cys Phe Thr Leu Phe Leu Ala Tyr Asp Thr Gln Leu Val

260 265 270

Leu Gly Asn Arg Lys His Thr Ile Ser Pro Glu Asp Tyr Ile Thr Gly

275 280 285

Ala Leu Gln Ile Tyr Thr Asp Ile Ile Tyr Ile Phe Thr Phe Val Leu
290 295 300

Gln Leu Met Gly Asp Arg Asn
305 310

[0052]

SEQ ID NO: 3

LENGTH: 383

TYPE: Amino acid

TOPOLOGY: Linear

MOLECULE TYPE: Protein

HYPOTHETICAL: No

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Lymphoma

CELL LINE: U937

CLONE: HP10493

SEQUENCE DESCRIPTION:

Met Ala Gly Ile Pro Gly Leu Leu Phe Leu Leu Phe Phe Leu Leu Cys
1 5 10 15

Ala Val Gly Gln Val Ser Pro Tyr Ser Ala Pro Trp Lys Pro Thr Trp
20 25 30

Pro Ala Tyr Arg Leu Pro Val Val Leu Pro Gln Ser Thr Leu Asn Leu
35 40 45

Ala Lys Pro Asp Phe Gly Ala Glu Ala Lys Leu Glu Val Ser Ser Ser
50 55 60

Cys Gly Pro Gln Cys His Lys Gly Thr Pro Leu Pro Thr Tyr Glu Glu
65 70 75 80

Ala Lys Gln Tyr Leu Ser Tyr Glu Thr Leu Tyr Ala Asn Gly Ser Arg
85 90 95

Thr Glu Thr Gln Val Gly Ile Tyr Ile Leu Ser Ser Ser Gly Asp Gly
100 105 110

Ala Gln His Arg Asp Ser Gly Ser Ser Gly Lys Ser Arg Arg Lys Arg
115 120 125

Gln Ile Tyr Gly Tyr Asp Ser Arg Phe Ser Ile Phe Gly Lys Asp Phe
130 135 140

Leu Leu Asn Tyr Pro Phe Ser Thr Ser Val Lys Leu Ser Thr Gly Cys
145 150 155 160

Thr Gly Thr Leu Val Ala Glu Lys His Val Leu Thr Ala Ala His Cys
165 170 175

Ile His Asp Gly Lys Thr Tyr Val Lys Gly Thr Gln Lys Leu Arg Val
180 185 190

Gly Phe Leu Lys Pro Lys Phe Lys Asp Gly Gly Arg Gly Ala Asn Asp
195 200 205

Ser Thr Ser Ala Met Pro Glu Gln Met Lys Phe Gln Trp Ile Arg Val
210 215 220

Lys Arg Thr His Val Pro Lys Gly Trp Ile Lys Gly Asn Ala Asn Asp
225 230 235 240

Ile Gly Met Asp Tyr Asp Tyr Ala Leu Leu Glu Leu Lys Lys Pro His

245 250 255

Lys Arg Lys Phe Met Lys Ile Gly Val Ser Pro Pro Ala Lys Gln Leu

260 265 270

Pro Gly Gly Arg Ile His Phe Ser Gly Tyr Asp Asn Asp Arg Pro Gly

275 280 285

Asn Leu Val Tyr Arg Phe Cys Asp Val Lys Asp Glu Thr Tyr Asp Leu

290 295 300

Leu Tyr Gln Gln Cys Asp Ala Gln Pro Gly Ala Ser Gly Ser Gly Val

305 310 315 320

Tyr Val Arg Met Trp Lys Arg Gln Gln Gln Lys Trp Glu Arg Lys Ile

325 330 335

Ile Gly Ile Phe Ser Gly His Gln Trp Val Asp Met Asn Gly Ser Pro

340 345 350

Gln Asp Phe Asn Val Ala Val Arg Ile Thr Pro Leu Lys Tyr Ala Gln

355 360 365

Ile Cys Tyr Trp Ile Lys Gly Asn Tyr Leu Asp Cys Arg Glu Gly

370 375 380

[0053]

SEQ ID NO: 4

LENGTH: 807

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Stomach cancer

CLONE: HP01207

SEQUENCE DESCRIPTION:

ATGGGCCAGA ACGACCTGAT GGGCACGGCC GAGGACTTCG CCGACCAGTT CCTCCGTGTC 60

ACAAAGCAGT ACCTGCCCCA CGTGGCGCGC CTCTGTCTGA TCAGCACCTT CCTGGAGGAC 120

GGCATCCGTA TGTGGTTCCA GTGGAGCGAG CAGCGCGACT ACATCGACAC CACCTGGAAC 180

TGCGGCTACC TGCTGGCCTC GTCCTTCGTC TTCCTCAACT TGCTGGGACA GCTGACTGGC 240

TGCGTCCTGG TGTTGAGCAG GAACTTCGTG CAGTACGCCT GCTTCGGGCT CTTTGGAATC 300

ATAGCTCTGC AGACGATTGC CTACAGCATT TTATGGGACT TGAAGTTTTT GATGAGGAAC 360

CTGGCCCTGG GAGGAGGCCT GTTGCTGCTC CTAGCAGAAT CCCGTTCTGA AGGGAAGAGC 420

ATGTTTGCGG GCGTCCCCAC CATGCGTGAG AGCTCCCCCA AACAGTACAT GCAGCTCGGA 480

GGCAGGGTCT TGCTGGTTCT GATGTTCATG ACCCTCCTTC ACTTTGACGC CAGCTTCTTT 540

TCTATTGTCC AGAACATCGT GGGCACAGCT CTGATGATTT TAGTGGCCAT TGGTTTTAAA 600

ACCAAGCTGG CTGCTTTGAC TCTTGTTGTG TGGCTCTTTG CCATCAACGT ATATTTCAAC 660

GCCTTCTGGA CCATTCCAGT CTACAAGCCC ATGCATGACT TCCTGAAATA CGACTTCTTC 720

CAGACCATGT CGGTGATTGG GGGCTTGCTC CTGGTGGTGG CCCTGGGCCC TGGGGGTGTC 780

TCCATGGATG AGAAGAAGAA GGAGTGG 807

[0054]

SEQ ID NO: 5

LENGTH: 933

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Stomach cancer

CLONE: HP01862

SEQUENCE DESCRIPTION:

ATGTCCAACC CCAGCGCCCC ACCACCATAT GAAGACCGCA ACCCCCTGTA CCCACGCCCT 60

CTGCCCCCTG GGGGCTATGG GCAGCCATCT GTCCTGCCAG GAGGGTATCC TGCCTACCCT 120

GGCTACCCGC AGCCTGGCTA CGGTCACCCT GCTGGCTACC CACAGCCCAT GCCCCCCACC 180

CACCCGATGC CCATGAACTA CGGCCCAGGC CATGGCTATG ATGGGGAGGA GAGAGCGGTG 240

AGTGATAGCT TCGGGCCTGG AGAGTGGGAT GACCGGAAAG TGCGACACAC TTTTATCCGA 300

AAGGTTTACT CCATCATCTC CGTGCAGCTG CTCATCACTG TGGCCATCAT TGCTATCTTC 360

ACCTTTGTGG AACCTGTCAG CGCCTTTGTG AGGAGAAATG TGGCTGTCTA CTACGTGTCC 420

TATGCTGTCT TCGTTGTCAC CTACCTGATC CTTGCCTGCT GCCAGGGACC CAGACGCCGT 480

TTCCCATGGA ACATCATTCT GCTGACCCTT TTTACTTTTG CCATGGGCTT CATGACGGGC 540

ACCATTTCCA GTATGTACCA AACCAAAGCC GTCATCATTG CAATGATCAT CACTGCGGTG 600

GTATCCATTT CAGTCACCAT CTTCTGCTTT CAGACCAAGG TGGACTTCAC CTCGTGCACA 660

GGCCTCTTCT GTGTCCTGGG AATTGTGCTC CTGGTGACTG GGATTGTCAC TAGCATTGTG 720

CTCTACTTCC AATACGTTTA CTGGCTCCAC ATGCTCTATG CTGCTCTGGG GGCCATTTGT 780

TTCACCCTGT TCCTGGCTTA CGACACACAG CTGGTCCTGG GGAACCGGAA GCACACCATC 840

AGCCCCGAGG ACTACATCAC TGGCGCCCTG CAGATTTACA CAGACATCAT CTACATCTTC 900

ACCTTTGTGC TGCAGCTGAT GGGGGATCGC AAT 933

[0055]

SEQ ID NO: 6

LENGTH: 1149

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Lymphoma

CELL LINE: U937

CLONE: HP10493

SEQUENCE DESCRIPTION:

ATGGCAGGGA TTCCAGGGCT CCTCTTCCTT CTCTTCTTTC TGCTCTGTGC TGTTGGGCAA 60

GTGAGCCCTT ACAGTGCCCC CTGGAAACCC ACTTGGCCTG CATACCGCCT CCCTGTCGTC 120

TTGCCCCAGT CTACCCTCAA TTTAGCCAAG CCAGACTTTG GAGCCGAAGC CAAATTAGAA 180

GTATCTTCTT CATGTGGACC CCAGTGTCAT AAGGGAACTC CACTGCCCAC TTACGAAGAG 240

GCCAAGCAAT ATCTGTCTTA TGAAACGCTC TATGCCAATG GCAGCCGCAC AGAGACGCAG 300

GTGGGCATCT ACATCCTCAG CAGTAGTGGA GATGGGGCCC AACACCGAGA CTCAGGGTCT 360

TCAGGAAAGT CTCGAAGGAA GCGGCAGATT TATGGCTATG ACAGCAGGTT CAGCATTTTT 420

CGGAAGGACT TCCTGCTCAA CTACCCTTTC TCAACATCAG TGAAGTTATC CACGGGCTGC 480

ACCGGCACCC TGGTGGCAGA GAAGCATGTC CTCACAGCTG CCCACTGCAT ACACGATGGA 540

AAAACCTATG TGAAAGGAAC CCAGAAGCTT CGAGTGGGCT TCCTAAAGCC CAAGTTTAAA 600

GATGGTGGTC GAGGGGCCAA CGACTCCACT TCAGCCATGC CCGAGCAGAT GAAATTTCAG 660

TGGATCCGGG TGAAACGCAC CCATGTGCCC AAGGGTTGGA TCAAGGGCAA TGCCAATGAC 720

ATCGGCATGG ATTATGATTA TGCCCTCCTG GAACTCAAAA AGCCCCACAA GAGAAAATTT 780

ATGAAGATTG GGGTGAGCCC TCCTGCTAAG CAGCTGCCAG GGGGCAGAAT TCACTTCTCT 840

GGTTATGACA ATGACCGACC AGGCAATTTG GTGTATCGCT TCTGTGACGT CAAAGACGAG 900

ACCTATGACT TGCTCTACCA GCAATGCGAT GCCCAGCCAG GGGCCAGCGG GTCTGGGGTC 960

TATGTGAGGA TGTGGAAGAG ACAGCAGCAG AAGTGGGAGC GAAAAATTAT TGGCATTTTT 1020

TCAGGGCACC AGTGGGTGGA CATGAATGGT TCCCCACAGG ATTTCAACGT GGCTGTCAGA 1080

ATCACTCCTC TCAAATATGC CCAGATTTGC TATTGGATTA AAGGAAACTA CCTGGATTGT 1140

AGGGAGGGG 1149

SEQ ID NO: 7

LENGTH: 2938

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Stomach cancer

CLONE: HP01207

FEATURES:

NAME/KEY: CDS

LOCATION: 101..910

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

AAAAAGGGCA CTTCCTGTGG AGGCCGCAGC GGGTGCGGGC GCCGACGGGC GAGAGCCAGC 60

GAGCGAGCGA GCGAGCCGAG CCGAGCCTCC CGCCGTCGCC ATG GGC CAG AAC GAC 115

Met Gly Gln Asn Asp

1 5

CTG ATG GGC ACG GCC GAG GAC TTC GCC GAC CAG TTC CTC CGT GTC ACA 163

Leu Met Gly Thr Ala Glu Asp Phe Ala Asp Gln Phe Leu Arg Val Thr

10 15 20

AAG CAG TAC CTG CCC CAC GTG GCG CGC CTC TGT CTG ATC AGC ACC TTC 211

Lys Gln Tyr Leu Pro His Val Ala Arg Leu Cys Leu Ile Ser Thr Phe

25 30 35

CTG GAG GAC GGC ATC CGT ATG TGG TTC CAG TGG AGC GAG CAG CGC GAC 259

Leu Glu Asp Gly Ile Arg Met Trp Phe Gln Trp Ser Glu Gln Arg Asp

40 45 50

TAC ATC GAC ACC ACC TGG AAC TGC GGC TAC CTG CTG GCC TCG TCC TTC 307

Tyr Ile Asp Thr Thr Trp Asn Cys Gly Tyr Leu Leu Ala Ser Ser Phe

55 60 65

GTC TTC CTC AAC TTG CTG GGA CAG CTG ACT GGC TGC GTC CTG GTG TTG 355

Val Phe Leu Asn Leu Leu Gly Gln Leu Thr Gly Cys Val Leu Val Leu

70 75 80 85

AGC AGG AAC TTC GTG CAG TAC GCC TGC TTC GGG CTC TTT GGA ATC ATA 403

Ser Arg Asn Phe Val Gln Tyr Ala Cys Phe Gly Leu Phe Gly Ile Ile

90 95 100

GCT CTG CAG ACG ATT GCC TAC AGC ATT TTA TGG GAC TTG AAG TTT TTG 451

Ala Leu Gln Thr Ile Ala Tyr Ser Ile Leu Trp Asp Leu Lys Phe Leu

105 110 115

ATG AGG AAC CTG GCC CTG GGA GGA GGC CTG TTG CTG CTC CTA GCA GAA 499

Met Arg Asn Leu Ala Leu Gly Gly Gly Leu Leu Leu Leu Leu Ala Glu

120 125 130

TCC CGT TCT GAA GGG AAG AGC ATG TTT GCG GGC GTC CCC ACC ATG CGT 547

Ser Arg Ser Glu Gly Lys Ser Met Phe Ala Gly Val Pro Thr Met Arg

135 140 145

GAG AGC TCC CCC AAA CAG TAC ATG CAG CTC GGA GGC AGG GTC TTG CTG 595

Glu Ser Ser Pro Lys Gln Tyr Met Gln Leu Gly Gly Arg Val Leu Leu

150 155 160 165

GTT CTG ATG TTC ATG ACC CTC CTT CAC TTT GAC GCC AGC TTC TTT TCT 643

Val Leu Met Phe Met Thr Leu Leu His Phe Asp Ala Ser Phe Phe Ser

170 175 180

ATT GTC CAG AAC ATC GTG GGC ACA GCT CTG ATG ATT TTA GTG GCC ATT 691

Ile Val Gln Asn Ile Val Gly Thr Ala Leu Met Ile Leu Val Ala Ile

185 190 195

GGT TTT AAA ACC AAG CTG GCT GCT TTG ACT CTT GTT GTG TGG CTC TTT 739

Gly Phe Lys Thr Lys Leu Ala Ala Leu Thr Leu Val Val Trp Leu Phe

200 205 210

GCC ATC AAC GTA TAT TTC AAC GCC TTC TGG ACC ATT CCA GTC TAC AAG 787

Ala Ile Asn Val Tyr Phe Asn Ala Phe Trp Thr Ile Pro Val Tyr Lys

215 220 225

CCC ATG CAT GAC TTC CTG AAA TAC GAC TTC TTC CAG ACC ATG TCG GTG 835

Pro Met His Asp Phe Leu Lys Tyr Asp Phe Phe Gln Thr Met Ser Val

230 235 240 245

ATT GGG GGC TTG CTC CTG GTG GTG GCC CTG GGC CCT GGG GGT GTC TCC 883

Ile Gly Gly Leu Leu Leu Val Val Ala Leu Gly Pro Gly Gly Val Ser

250 255 260

ATG GAT GAG AAG AAG AAG GAG TGG TAA CAGTCACAGA TCCCTACCTG 930

Met Asp Glu Lys Lys Lys Glu Trp

265

CCTGGCTAAG ACCCGTGGCC GTCAAGGACT GGTTCGGGGT GGATTCAACA AAACTGCCAG 990

CTTTTATGTA TCCTCTTCCC TTCCCCTCCC TTGGTAAAGG CACAGATGTT TTGAGAACTT 1050

TATTTGCAGA GACACCTGAG AATCGATGGC TCAGTCTGCT CTGGAGCCAC AGTCTGGCGT 1110

CTGACCCTTC AGTGCAGGCC AGCCTGGCAG CTGGAAGCCT CCCCCACGCC GAGGCTTTGG 1170

AGTGAACAGC CCGCTTGGCT GTGGCATCTC AGTCCTATTT TTGAGTTTTT TTGTGGGGGT 1230

ACAGGAGGGG GCCTTCAAGC TGTACTGTGA GCAGACGCAT TGGTATTATC ATTCAAAGCA 1290

GTCTCCCTCT TATTTGTAAG TTTACATTTT TAGCGGAAAC TACTAAATTA TTTTGGGTGG 1350

TTCAGCCAAA CCTCAAAACA GTTAATCTCC CTGGTTTAAA ATCACACCAG TGGCTTTGAT 1410

GTTGTTTCTG CCCCGCATTG TATTTTATAG GAATACTGAA AACATTTAGG GACACCCAAA 1470

GAATGATGCA GTATTAAAGG GGTGGTAGAA GCTGCTGTTT ATGATAAAAG TCATCGGTCA 1530

GAAAATCAGC TTGGATTGGT GCCAAGTGTT TTATTGGGTA ACACCCTGGG AGTTTTAGTA 1590

GCTTGAGGCA AGGTGGAGGG GCAAGAAGTC CTTGGGGAAG CTGCTGGTCT GGGTGCTGCT 1650

GGCCTCCAAG CTGGCAGTGC GAAGGGCTAG TGAGACCACA CAGGGGTAGC CCCAGCAGCA 1710

GCACCCTGCA AGCCAGCCTG GCCAGCTGCT CAGACCAGCT TGCAGAGCCG CAGCCGCTGT 1770

GGGCAGGGGG TGTGGCAGGA GCTCCCAGCA CTGGAGACCC ACGGACTCAA CCCAGTTACC 1830

TCACATGGGG CCTTTTCTGA GCAAGGTCTC GAAAGCGCAG GCCGCCCTGG CTGAGCAGCA 1890

CCGCCCTTTC CCAGCTGCAC TCGCCCTGTG GACAGCCCCG ACACACCACT TTCCTGAGGC 1950

TGTCGCTCAC TCAGATTGTC CGTTTGCTAT GCCGAATGCA GCCAAAATTC CTTTTTACAA 2010

TTTGTGATGC CTTACCGATT TGATCTTAAT CCTGTATTTA AAGTTTTCTA ACACTGCCTT 2070

ATACTGTGTT TCTCTTTTTG GGGGAGCTTA ACTGCTTGTT GCTCCCTGTC GTCTGCACCA 2130

TAGTAAATGC CACAAGGGTA GTCGAACACC TCTCTGGCCC CTAGACCTAT CTGGGGACAG 2190

GCTGGCTCAG CCTGTCTCCA GGGCTGCTGC GGCCCAGCCC CGAGCCTGCC TCCCTCTTGG 2250

CCTCTCATCC ATTGGCTCTG CAGGGCAGGG GTGAGGCAGG TTTCTGCTCA TAAGTGCTTT 2310

TGGAAGTCAC CTACCTTTTT AACACAGCCG AACTAGTCCC AACGCGTTTG CAAATATTCC 2370

CCTGGTAGCC TACTTCCTTA CCCCCGAATA TTGGTAAGAT CGATCAATGG CTTCAGGACA 2430

TGGGTTCTCT TCTCCTGTGA TCATTCAAGT GCTCACTGCA TGAAGACTGG CTTGTCTCAG 2490

TGTTTCAACC TCACCAGGGC TGTCTCTTGG TCCACACCTC GCTCCCTGTT AGTGCCGTAT 2550

GACAGCCCCC ATCAAATGAC CTTGGCCAAG TCACGGTTTC TCTGTGGTCA AGGTTGGTTG 2610

GCTGATTGGT GGAAAGTAGG GTGGACCAAA GGAGGCCACG TGAGCAGTCA GCACCAGTTC 2670

TGCACCAGCA GCGCCTCCGT CCTAGTGGGT GTTCCTGTTT CTCCTGGCCC TGGGTGGGCT 2730

AGGGCCTGAT TCGGGAAGAT GCCTTTGCAG GGAGGGGAGG ATAAGTGGGA TCTACCAATT 2790

GATTCTGGCA AAACAATTTC TAAGATTTTT TTGCTTTATG TGGGAAACAG ATCTAAATCT 2850

CATTTTATGC TGTATTTTAT ATCTTAGTTG TGTTTGAAAA CGTTTTGATT TTTGGAAACA 2910

CATCAAAATA AATAATGGCG TTTGTTGT 2938

[0056]

SEQ ID NO: 8

LENGTH: 2290

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL TYPE: Stomach cancer

CLONE: HP01862

FEATURES:

NAME/KEY: CDS

LOCATION: 81..1016

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

ACACTCCGAG GCCAGGAACG CTCCGTCTGG AACGGCGCAG GTCCCAGCAG CTGGGGTTCC 60

CCCTCAGCCC GTGAGCAGCC ATG TCC AAC CCC AGC GCC CCA CCA CCA TAT GAA 113

Met Ser Asn Pro Ser Ala Pro Pro Pro Tyr Glu

1 5 10

GAC CGC AAC CCC CTG TAC CCA GGC CCT CTG CCC CCT GGG GGC TAT GGG 161

Asp Arg Asn Pro Leu Tyr Pro Gly Pro Leu Pro Pro Gly Gly Tyr Gly

15 20 25

CAG CCA TCT GTC CTG CCA GGA GGG TAT CCT GCC TAC CCT GGC TAC CCG 209

Gln Pro Ser Val Leu Pro Gly Gly Tyr Pro Ala Tyr Pro Gly Tyr Pro

30 35 40

CAG CCT GGC TAC GGT CAC CCT GCT GGC TAC CCA CAG CCC ATG CCC CCC 257

Gln Pro Gly Tyr Gly His Pro Ala Gly Tyr Pro Gln Pro Met Pro Pro

45 50 55

ACC CAC CCG ATG CCC ATG AAC TAC GGC CCA GGC CAT GGC TAT GAT GGG 305

Thr His Pro Met Pro Met Asn Tyr Gly Pro Gly His Gly Tyr Asp Gly

60 65 70 75

GAG GAG AGA GCG GTG AGT GAT AGC TTC GGG CCT GGA GAG TGG GAT GAC 353

Glu Glu Arg Ala Val Ser Asp Ser Phe Gly Pro Gly Glu Trp Asp Asp

80 85 90

CGG AAA GTG CGA CAC ACT TTT ATC CGA AAG GTT TAC TCC ATC ATC TCC 401

Arg Lys Val Arg His Thr Phe Ile Arg Lys Val Tyr Ser Ile Ile Ser

95 100 105

GTG CAG CTG CTC ATC ACT GTG GCC ATC ATT GCT ATC TTC ACC TTT GTG 449

Val Gln Leu Leu Ile Thr Val Ala Ile Ile Ala Ile Phe Thr Phe Val

110 115 120

GAA CCT GTC AGC GCC TTT GTG AGG AGA AAT GTG GCT GTC TAC TAC GTG 497

Glu Pro Val Ser Ala Phe Val Arg Arg Asn Val Ala Val Tyr Tyr Val

125 130 135

TCC TAT GCT GTC TTC GTT GTC ACC TAC CTG ATC CTT GCC TGC TGC CAG 545

Ser Tyr Ala Val Phe Val Val Thr Tyr Leu Ile Leu Ala Cys Cys Gln

140 145 150 155

GGA CCC AGA CGC CGT TTC CCA TGG AAC ATC ATT CTG CTG ACC CTT TTT 593

Gly Pro Arg Arg Arg Phe Pro Trp Asn Ile Ile Leu Leu Thr Leu Phe

160 165 170

ACT TTT GCC ATG GGC TTC ATG ACG GGC ACC ATT TCC AGT ATG TAC CAA 641

Thr Phe Ala Met Gly Phe Met Thr Gly Thr Ile Ser Ser Met Tyr Gln

175 180 185

ACC AAA GCC GTC ATC ATT GCA ATG ATC ATC ACT GCG GTG GTA TCC ATT 689
Thr Lys Ala Val Ile Ile Ala Met Ile Ile Thr Ala Val Val Ser Ile
190 195 200

TCA GTC ACC ATC TTC TGC TTT CAG ACC AAG GTG GAC TTC ACC TCG TGC 737
Ser Val Thr Ile Phe Cys Phe Gln Thr Lys Val Asp Phe Thr Ser Cys
205 210 215

ACA GGC CTC TTC TGT GTC CTG GGA ATT GTG CTC CTG GTG ACT GGG ATT 785
Thr Gly Leu Phe Cys Val Leu Gly Ile Val Leu Leu Val Thr Gly Ile
220 225 230 235

GTC ACT AGC ATT GTG CTC TAC TTC CAA TAC GTT TAC TGG CTC CAC ATG 833
Val Thr Ser Ile Val Leu Tyr Phe Gln Tyr Val Tyr Trp Leu His Met
240 245 250

CTC TAT GCT GCT CTG GGG GCC ATT TGT TTC ACC CTG TTC CTG GCT TAC 881
Leu Tyr Ala Ala Leu Gly Ala Ile Cys Phe Thr Leu Phe Leu Ala Tyr
255 260 265

GAC ACA CAG CTG GTC CTG GGG AAC CGG AAG CAC ACC ATC AGC CCC GAG 929
Asp Thr Gln Leu Val Leu Gly Asn Arg Lys His Thr Ile Ser Pro Glu
270 275 280

GAC TAC ATC ACT GGC GCC CTG CAG ATT TAC ACA GAC ATC ATC TAC ATC 977
Asp Tyr Ile Thr Gly Ala Leu Gln Ile Tyr Thr Asp Ile Ile Tyr Ile
285 290 295

TTC ACC TTT GTG CTG CAG CTG ATG GGG GAT CGC AAT TAAGGAG 1020
Phe Thr Phe Val Leu Gln Leu Met Gly Asp Arg Asn
300 305 310

CAAGCCCCCA TTTTCACCCG ATCCTGGGCT CTCCCTTCCA AGCTAGAGGG CTGGGCCCTA 1080

TGACTGTGGT CTGGGCTTTA GGCCCCTTTC CTTCCCCTTG AGTAACATGC CCAGTTTCCT 1140

TTCTGTCCTG GAGACAGGTG GCCTCTCTGG CTATGGATGT GTGGGTACTT GGTGGGGACG 1200

GAGGAGCTAG GGACTAACTG TTGCTCTTGG TGGGCTTGGC AGGGACTAGG CTGAAGATGT 1260

GTCTTCTCCC CGCCACCTAC TGTATGACAC CACATTCTTC CTAACAGCTG GGGTTGTGAG 1320

GAATATGAAA AGAGCCTATT CGATAGCTAG AAGGGAATAT GAAAGGTAGA AGTGACTTCA 1380

AGGTCACGAG GTTCCCCTCC CACCTCTGTC ACAGGCTTCT TGACTACGTA GTTGGAGCTA 1440

TTTCTTCCCC CAGCAAAGCC AGAGAGCTTT GTCCCCGGCC TCCTGGACAC ATAGGCCATT 1500

ATCCTGTATT CCTTTGGCTT GGCATCTTTT AGCTCAGGAA GGTAGAAGAG ATCTGTGCCC 1560

ATGGGTCTCC TTGCTTCAAT CCCTTCTTGT TTCAGTGACA TATGTATTGT TTATCTGGGT 1620

TAGGGATGGG GGACAGATAA TAGAACGAGC AAAGTAACCT ATACAGGCCA GCATGGAACA 1680

GCATCTCCCC TGGGCTTGCT CCTGGCTTGT GACGCTATAA GACAGAGCAG GCCACATGTG 1740

GCCATCTGCT CCCCATTCTT GAAAGCTGCT GGGGCCTCCT TGCAGGCTTC TGGATCTCTG 1800

GTCAGAGTGA ACTCTTGCTT CCTGTATTCA GGCAGCTCAG AGCAGAAAGT AAGGGGCAGA 1860

GTCATACGTG TGGCCAGGAA GTAGCCAGGG TGAAGAGAGA CTCGGTGCGG GCAGGGAGAA 1920

TGCCTGGGGG TCCCTCACCT GGCTAGGGAG ATACCGAAGC CTACTGTGGT ACTGAAGACT 1980

TCTGGGTTCT TTCCTTCTGC TAACCCAGGG AGGGTCCTAA GAGGAAGGTG ACTTCTCTCT 2040

GTTTGTCTTA AGTTGCACTG GGGGATTTCT GACTTGAGGC CCATCTCTCC AGCCAGCCAC 2100

TGCCTTCTTT GTAATATTAA GTGCCTTGAG CTGGAATGGG GAAGGGGGAC AAGGGTCAGT 2160

CTGTCGGGTG GGGGCAGAAA TCAAATCAGC CCAAGGATAT AGTTAGGATT AATTACTTAA 2220

TAGAGAAATC CTAACTATAT CACACAAAGG GATACAACTA TAAATGTAAT AAAATTTATG 2280

TCTAGAAGTT 2290

[0057]

SEQ ID NO: 9

LENGTH: 3705

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: *Homo sapiens*

CELL LINE: U937

CLONE: HP10493

FEATURES:

NAME/KEY: CDS

LOCATION: 124..1275

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

```
ACTCTCGGCT GTGCGGCGGG GCAGGCATGG GAGCCGCGCG CTCTCTCCCG GCGCCCACAC      60
CTGTCTGAGC GGCGCAGCGA GCCGCGGCCC GGGCGGGCTG CTCGGCGCGG AACAGTGCTC     120
GGC ATG GCA GGG ATT CCA GGG CTC CTC TTC CTT CTC TTC TTT CTG CTC        168
   Met Ala Gly Ile Pro Gly Leu Leu Phe Leu Leu Phe Phe Leu Leu
     1               5                  10                  15
TGT GCT GTT GGG CAA GTG AGC CCT TAC AGT GCC CCC TGG AAA CCC ACT        216
Cys Ala Val Gly Gln Val Ser Pro Tyr Ser Ala Pro Trp Lys Pro Thr
                20                  25                  30
TGG CCT GCA TAC CGC CTC CCT GTC GTC TTG CCC CAG TCT ACC CTC AAT        264
```

Trp Pro Ala Tyr Arg Leu Pro Val Val Leu Pro Gln Ser Thr Leu Asn

35 40 45

TTA GCC AAG CCA GAC TTT GGA GCC GAA GCC AAA TTA GAA GTA TCT TCT 312

Leu Ala Lys Pro Asp Phe Gly Ala Glu Ala Lys Leu Glu Val Ser Ser

50 55 60

TCA TGT GGA CCC CAG TGT CAT AAG GGA ACT CCA CTG CCC ACT TAC GAA 360

Ser Cys Gly Pro Gln Cys His Lys Gly Thr Pro Leu Pro Thr Tyr Glu

65 70 75

GAG GCC AAG CAA TAT CTG TCT TAT GAA ACG CTC TAT GCC AAT GGC AGC 408

Glu Ala Lys Gln Tyr Leu Ser Tyr Glu Thr Leu Tyr Ala Asn Gly Ser

80 85 90 95

CGC ACA GAG ACG CAG GTG GGC ATC TAC ATC CTC AGC AGT AGT GGA GAT 456

Arg Thr Glu Thr Gln Val Gly Ile Tyr Ile Leu Ser Ser Ser Gly Asp

100 105 110

GGG GCC CAA CAC CGA GAC TCA GGG TCT TCA GGA AAG TCT CGA AGG AAG 504

Gly Ala Gln His Arg Asp Ser Gly Ser Ser Gly Lys Ser Arg Arg Lys

115 120 125

CGG CAG ATT TAT GGC TAT GAC AGC AGG TTC AGC ATT TTT GGG AAG GAC 552

Arg Gln Ile Tyr Gly Tyr Asp Ser Arg Phe Ser Ile Phe Gly Lys Asp

130 135 140

TTC CTG CTC AAC TAC CCT TTC TCA ACA TCA GTG AAG TTA TCC ACG GGC 600

Phe Leu Leu Asn Tyr Pro Phe Ser Thr Ser Val Lys Leu Ser Thr Gly

145 150 155

TGC ACC GGC ACC CTG GTG GCA GAG AAG CAT GTC CTC ACA GCT GCC CAC 648

Cys Thr Gly Thr Leu Val Ala Glu Lys His Val Leu Thr Ala Ala His

160 165 170 175

TGC ATA CAC GAT GGA AAA ACC TAT GTG AAA GGA ACC CAG AAG CTT CGA 696
Cys Ile His Asp Gly Lys Thr Tyr Val Lys Gly Thr Gln Lys Leu Arg

180 185 190

GTG GGC TTC CTA AAG CCC AAG TTT AAA GAT GGT GGT CGA GGG GCC AAC 744
Val Gly Phe Leu Lys Pro Lys Phe Lys Asp Gly Gly Arg Gly Ala Asn

195 200 205

GAC TCC ACT TCA GCC ATG CCC GAG CAG ATG AAA TTT CAG TGG ATC CGG 792
Asp Ser Thr Ser Ala Met Pro Glu Gln Met Lys Phe Gln Trp Ile Arg

210 215 220

GTG AAA CGC ACC CAT GTG CCC AAG GGT TGG ATC AAG GGC AAT GCC AAT 840
Val Lys Arg Thr His Val Pro Lys Gly Trp Ile Lys Gly Asn Ala Asn

225 230 235

GAC ATC GGC ATG GAT TAT GAT TAT GCC CTC CTG GAA CTC AAA AAG CCC 888
Asp Ile Gly Met Asp Tyr Asp Tyr Ala Leu Leu Glu Leu Lys Lys Pro

240 245 250 255

CAC AAG AGA AAA TTT ATG AAG ATT GGG GTG AGC CCT CCT GCT AAG CAG 936
His Lys Arg Lys Phe Met Lys Ile Gly Val Ser Pro Pro Ala Lys Gln

260 265 270

CTG CCA GGG GGC AGA ATT CAC TTC TCT GGT TAT GAC AAT GAC CGA CCA 984
Leu Pro Gly Gly Arg Ile His Phe Ser Gly Tyr Asp Asn Asp Arg Pro

275 280 285

GGC AAT TTG GTG TAT CGC TTC TGT GAC GTC AAA GAC GAG ACC TAT GAC 1032
Gly Asn Leu Val Tyr Arg Phe Cys Asp Val Lys Asp Glu Thr Tyr Asp

290 295 300

TTG CTC TAC CAG CAA TGC GAT GCC CAG CCA GGG GCC AGC GGG TCT GGG 1080
Leu Leu Tyr Gln Gln Cys Asp Ala Gln Pro Gly Ala Ser Gly Ser Gly
305 310 315
GTC TAT GTG AGG ATG TGG AAG AGA CAG CAG CAG AAG TGG GAG CGA AAA 1128
Val Tyr Val Arg Met Trp Lys Arg Gln Gln Gln Lys Trp Glu Arg Lys
320 325 330 335
ATT ATT GGC ATT TTT TCA GGG CAC CAG TGG GTG GAC ATG AAT GGT TCC 1176
Ile Ile Gly Ile Phe Ser Gly His Gln Trp Val Asp Met Asn Gly Ser
340 345 350
CCA CAG GAT TTC AAC GTG GCT GTC AGA ATC ACT CCT CTC AAA TAT GCC 1224
Pro Gln Asp Phe Asn Val Ala Val Arg Ile Thr Pro Leu Lys Tyr Ala
355 360 365
CAG ATT TGC TAT TGG ATT AAA GGA AAC TAC CTG GAT TGT AGG GAG GGG 1272
Gln Ile Cys Tyr Trp Ile Lys Gly Asn Tyr Leu Asp Cys Arg Glu Gly
370 375 380
TGACACAG TGTTCCCTCC TGGCAGCAAT TAAGGGTCTT CATGTTCTTA TTTTAGGAGA 1330
GGCCAAATTG TTTTTTGTCA TTGGCGTGCA CACGTGTGTG TGTGTGTGTG TGTGTAAGGT 1390
GTCTTATAAT CTTTTACCTA TTTCTTACAA TTGCAAGATG ACTGGCTTTA CTATTTGAAA 1450
ACTGGTTTGT GTATCATATC ATATATCATT TAAGCAGTTT GAAGGCATAC TTTTGCATAG 1510
AAATAAAAAA AATACTGATT TGGGGCAATG AGGAATATTT GACAATTAAG TTAATCTTCA 1570
CGTTTTTGCA AACTTTGATT TTTATTTCAT CTGAACTTGT TTCAAAGATT TATATTAAAT 1630
ATTTGGCATA CAAGAGATAT GAATTCTTAT ATGTGTGCAT GTGTGTTTTC TTCTGAGATT 1690
CATCTTGGTG GTGGGTTTTT TTGTTTTTTT AATTCAGTGC CTGATCTTTA ATGCTTCCAT 1750
AAGGCAGTGT TCCCATTTAG GAACTTTGAC AGCATTTGTT AGGCAGAATA TTTTGGATTT 1810
GGAGGCATTT GCATGGTAGT CTTTGAACAG TAAAATGATG TGTTGACTAT ACTGATACAC 1870

ATATTAAACT ATACCTTATA GTAAACCAGT ATCCCAAGCT GCTTTTAGTT CCAAAAATAG 1930

TTTCTTTTCC AAAGGTTGTT GCTCTACTTT GTAGGAAGTC TTTGCATATG GCCCTCCCAA 1990

CTTTAAAGTC ATACCAGAGT GGCCAAGAGT GTTTATCCCA ACCCTTCCAT TTAACAGGAT 2050

TTCACTCACA TTTCTGGAAC TAGCTATTTT TCAGAAGACA ATAATCAGGG CTTAATTAGA 2110

ACAGGCTGTA TTTCCTCCCA GCAAACAGTT GTGGCCACAC TAAAAACAAT CATAGCATTT 2170

TACCCCTGGA TTATAGCACA TCTCATGTTT TATCATTTGG ATGGAGTAAT TTAAAATGAA 2230

TTAAATTCCA GAGAACAATG GAAGCATTGC CTGGCAGATG TCACAACAGA ATAACCACTT 2290

GTTTGGAGCC TGGCACAGTC CTCCAGCCTG ATCAAAAATT ATTCTGCATA GTTTTCAGTG 2350

TGCTTTCTGG GAGCTATGTA CTTCTTCAAT TTGGAAACTT TTCTCTCTCA TTTATAGTGA 2410

AAATACTTGG AAGTTACTTT AAGAAAACCA GTGTGGCCTT TTTCCCTCTA GCTTTAAAAG 2470

GGCCGCTTTT GCTGGAATGC TCTAGGTTAT AGATAAACAA TTAGGTATAA TAGCAAAAAT 2530

GAAAATTGGA AGAATGCAAA ATGGATCAGA ATCATGCCTT CCAATAAAGG CCTTTACACA 2590

TGTTTTATCA ATATGATTAT CAAATCACAG CATATACAGA AAAGACTTGG ACTTATTGTA 2650

TGTTTTTATT TTATGGCTCT CGGCCTAAGC ACTTCTTTCT AAATGTATCG GAGAAAAAAT 2710

CAAATGGACT ACAAGCACGT GTTTGCTGTG CTTGCACCCC AGGTAAACCT GCATTGTAGC 2770

AATTTGTAAG GATATTCAGA TGGAGCACTG TCACTTAGAC ATTCTCTGGG GGATTTTCTG 2830

CTTGTCTTTC TTGAGCTTTT TGGAAGGATA ATTCTGATAA GGCACTCAAG AAACGTACAA 2890

CCACAGTGCT TTCTTCAAAT CATATGAGAA ATACTATGCA TAGCAAGGAG ATGCAGAGCC 2950

GCCAGGAAAA TTCTGAGTTC CAGCACAATT TTCTTTGGAA TCTAACAGGA ATCTAGCCTG 3010

AGGAAGAAGG GAGGTCTCCA TTTCTATGTC TGGTATTTGG GGGTTTTGTT TGTTTTTGCT 3070

TTAGCTTGGT GAAAAAAAGT TCACTGAACA CCAAGACCAG AATGGATTTT TTTAAAAAAA 3130

TAGATGTTCC TTTTGTGAAG CACCTTGATT CCTTGATTTT GATTTTTTGC AAAGTTAGAC 3190

AATGGCACAA AGTCAAAATG AAATCAATGT TTAGTTCACA AGTAGATGTA ATTTACTAAA 3250

GAATGATACA CCCATATGCT ATATACAGCT TAACTCACAG AACTGTAAAA GAAAATTATA 3310

AAATAATTCA ACATGTCCAT CTTTTTAGTG ATAATAAAAG AAAGCATGGT ATTAAACTAT 3370

CATAGAAGTA GACAGAAAAA GAAAAAAGGA CTCATGGCAT TATTAATATA ATTAGTGCTT 3430

TACATGTGTT AGTTATACAT ATTAGAAGCA TATTTGCCTA GTAAGGCTAG TAGAACCACA 3490

TTTCCCAAAG TGTGCTCCTT AAACACTCAT GCCTTATGAT TTTCTACCAA AAGTAAAAAG 3550

GGTTGTATTA AGTCAGAGGA AGATGCCTCT CCATTTTCCC TCTCTTTATC AGAGGTTCAC 3610

ATGCCTGTCT GCACATTAAA AGCTCTGGGA AGACCTGTTG TAAAGGGACA AGTTGAGGTT 3670

GTAAAATCTG CATTTAAATA AACATCTTTG ATCAC 3705

[0058]

Brief Description of the Drawings:

Figure 1: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP01207.

Figure 2: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP01862.

Figure 3: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10493.

Document Name: Abstract

Abstract:

Problems to be Solved: To provide human proteins having transmembrane domains, cDNAs coding for these proteins, expression vectors of said cDNAs and eucaryotic cells expressing said cDNAs.

Means to Solve the Problems: Proteins containing any of the amino acid sequences represented by Sequence Nos. 1 to 3, DNAs coding for these proteins, exemplified by cDNAs containing any of the base sequences represented by Sequence Nos. 4 to 6, expression vectors of said cDNAs, as well as eucaryotic cells expressing said cDNAs. Said proteins and eucaryotic cells having said proteins on the surface of membrane can be provided by expressing cDNAs encoding human proteins having transmembrane domains and recombinants of these human cDNAs.

Selected Figure: None

Document Name: Drawings

Fig. 1

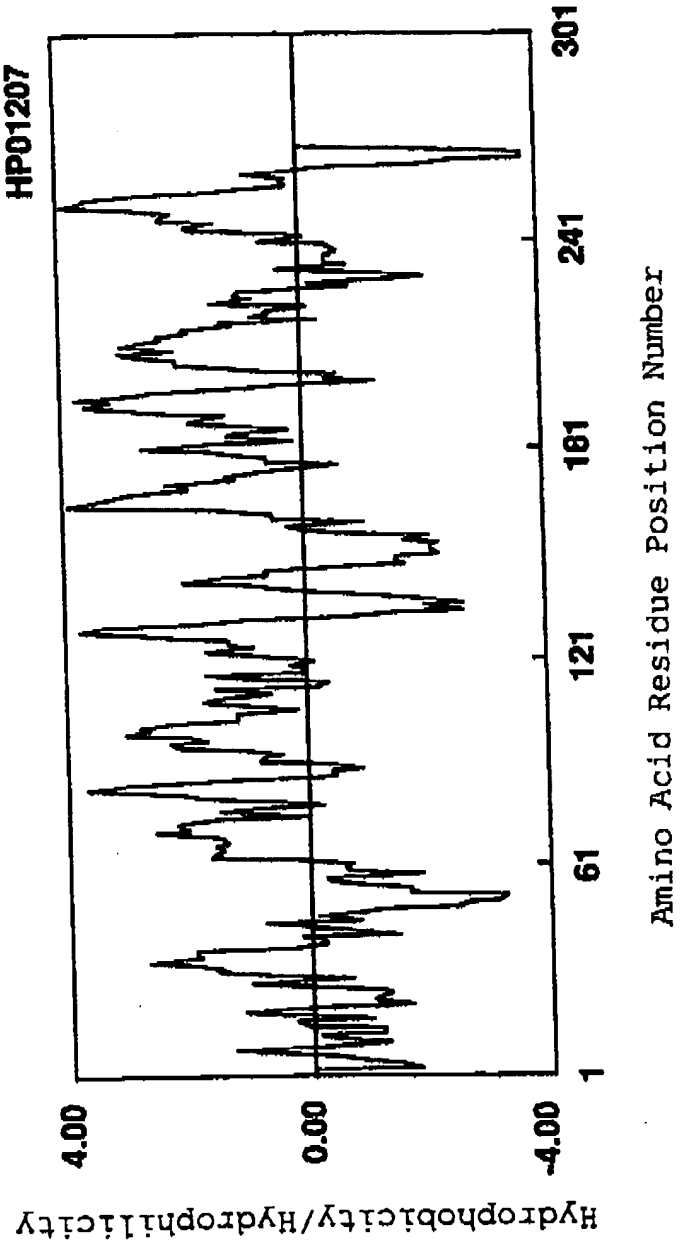


Fig. 2

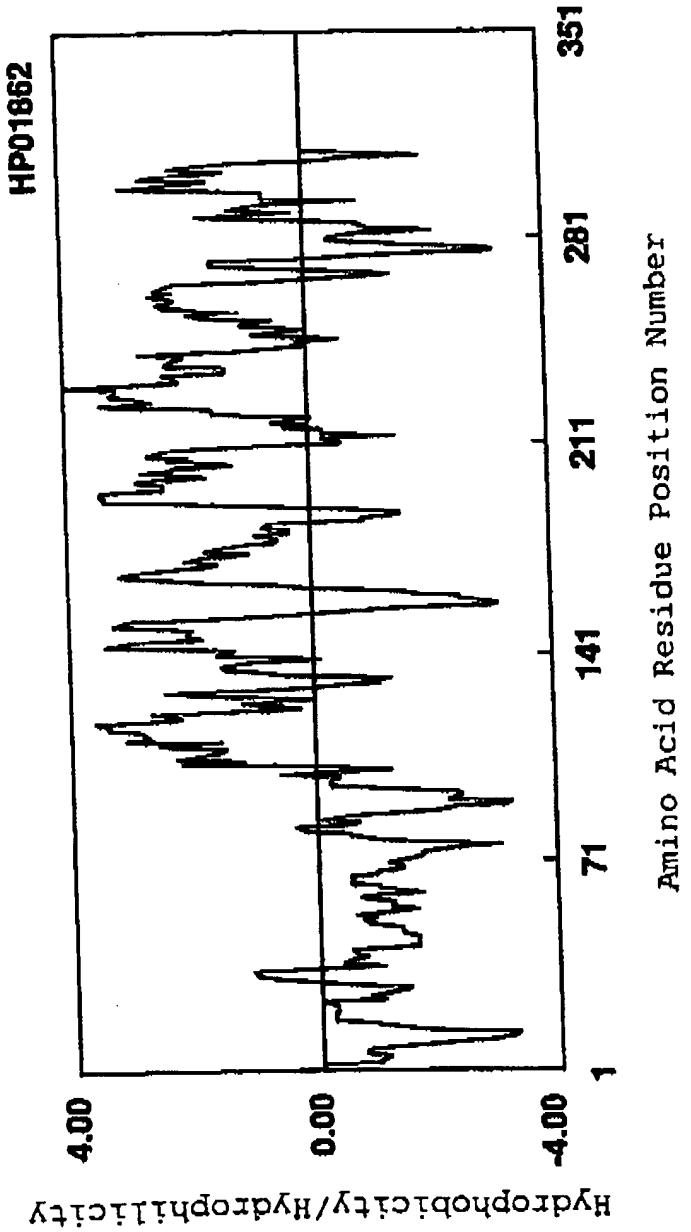


Fig. 3